

CHAPITRE 6 : Vers le sémantique : analyse thématique du genre

Si nous avons jusqu'à présent observé le genre dans ses régulations morphosyntaxiques, structurelles et configuratives, nous n'avons pas encore apprécié l'objet comme espace sémantiquement normé de conceptualisation et d'élaboration scientifique.

De manière générale, les textes scientifiques sont souvent réduits aux *concepts* ou aux *termes* (notions souvent employées de manière synonymique¹) qu'ils contiennent et véhiculent :

La question du texte scientifique reste difficile à poser : la tradition scientifique occidentale fait de la science une affaire de concepts et de termes, non de textes, car elle tient que l'objectivité est indépendante de la différence des langues et des normes textuelles. (Rastier, 2005)

et que l'on cherche à extraire et repérer pour constituer des dictionnaires sémantiques ou des ontologies, et faire de la recherche d'information (*e.g.* veille scientifique et stratégique, SanJuan & Ibekwe-SanJuan, 2002).

Il nous semble exclu, étant donné notre cadre général d'analyse et notre intérêt pour le genre et le texte, d'adopter une telle perspective : procéder à un étiquetage sémantique de notre corpus à l'aide d'ontologies comme Wordnet ou Eurowordnet, qui ne tiennent aucun compte des discours et des genres (Rastier, 2004), serait peu pertinent, et constituer un dictionnaire, ou un lexique des *termes*, ou plus concrètement, des *lexicalisations* les plus représentées et les plus également réparties dans le corpus occulte la dimension textuelle du genre et confronte le corpus à ses biais (impact des numéros thématiques sur la définition des concepts, incidence des revues sélectionnées et des spécificités de leurs objets, etc.). A fortiori, les dictionnaires des sciences du langage ne font pas défaut, et Gobert a publié en 2001 un glossaire bibliographique indexant les articles de quatre revues, d'ailleurs représentées dans notre corpus (Langage, Langue Française, HEL, LINX).

Ainsi, si les dictionnaires, les ontologies et les bases terminologiques sont en pleine expansion, la description des concepts scientifiques en tant qu'*unités textuelles et textualisées* est encore peu développée ; le *contexte* des termes ou des concepts est certes pris en compte par les dictionnaires contextuels, et le courant de la terminologie textuelle s'attache certes à objectiver la sélection des termes en corpus (par une analyse des spécificités par exemple, L'homme, 2004), mais peu d'entreprises cherchent à caractériser les concepts en tant qu'*unités méso-sémantiques* (palier intermédiaire entre la micro-sémantique du mot et la macro-sémantique du texte) ; dans cette perspective, les concepts, que l'on peut décrire comme des *thèmes*, sont potentiellement corrélés à des marqueurs ou à des formes expressives relevant de tous les niveaux de l'analyse linguistique (Rastier, 2005).

Cette conception thématique du concept scientifique a montré son intérêt et son efficacité dans diverses études récentes (Loiseau 2003, Valette 2003), et s'inscrit de manière très

¹ « Pour la théorie terminologique, les termes ne sont pas des signifiés mais des concepts ; elle confère donc à certains mots le statut de termes et en fait alors l'expression d'unités conceptuelles non liées à la langue. Inspirée par le positivisme logique, la doctrine terminologique de Wüster considérait ainsi les concepts comme des représentations indépendantes des langues et le cognitivisme perpétue aujourd'hui cette tradition. » (Rastier, *Ibid.*)

pertinente dans une conception multidimensionnelle du genre, au contraire des approches terminologiques et dictionnaires, qui demeurent souvent centrées sur le seul niveau lexical.

Le présent chapitre se veut exploratoire, et cherche à éprouver différentes pistes et différents critères typologiques potentiellement discriminants – on soulignera le caractère interactif et non isolé de ces six critères :

- La *fréquence* et la *répartition* des objets dans les textes ;
- Les *corrélations morphosyntaxiques* des candidats : nous avons ensuite examiné les corrélations morphosyntaxiques d'un sous-ensemble de candidats au singulier et au pluriel, afin de déterminer si ces derniers étaient éventuellement corrélés à des phénomènes linguistiques particuliers, voire à des dimensions textuelles ;
- *Configuration tactique* : on s'est également intéressé au positionnement des candidats dans la séquentialité du texte. Le genre de l'article étant structurellement normé (v. chapitre 4), la position du concept dans la séquentialité de l'article ne nous semble pas anodine : un concept discuté en début d'article a plus de chances d'être problématisé qu'un concept positionné en son corps ;
- *Co-occurents* : dans la mesure où ils participent à la qualification des thèmes, les co-occurents des candidats ont été observés afin d'apprécier leur(s) régulation(s) sémantique(s) ; on s'est ainsi intéressé au nombre de co-occurents, qui indique une plus ou moins grande stabilisation de la notion et aux éventuelles lexicalisations lieutenantes de la forme ;
- Incidence du *numéro thématique* : le corpus étant globalement partitionné en numéros thématiques de revue, nous avons cherché à éprouver la stabilité de certaines formes d'un numéro thématique à l'autre ;
- Incidence du *style d'auteur* : on a ensuite cherché à évaluer l'incidence du style d'auteur, et de son domaine scientifique d'appartenance – qui véhicule des conceptions et des présupposés distincts autour des mêmes formes – sur les candidats.

6.1. Examen des hautes fréquences

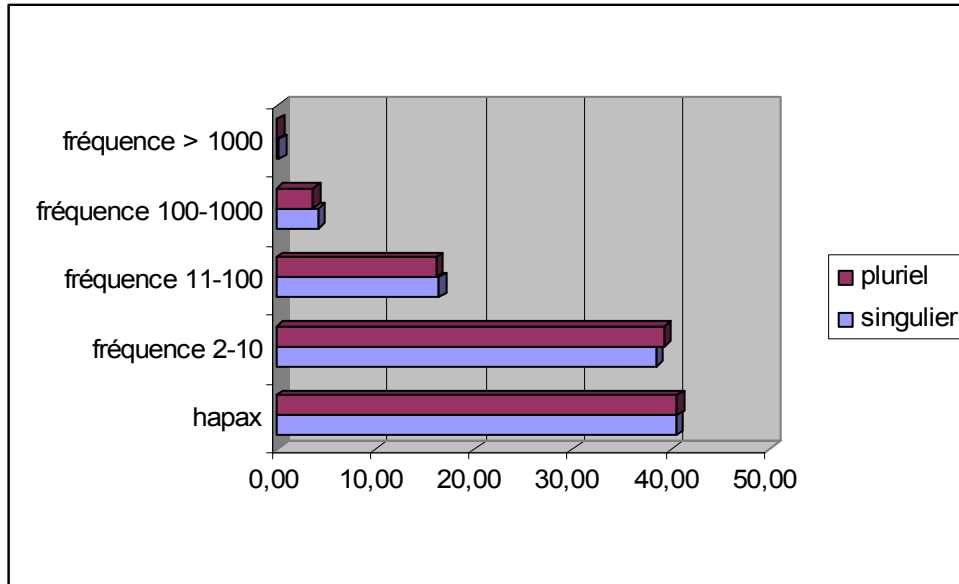
Supposés constituer un mode d'accès privilégié aux thèmes scientifiques linguistiques, ce sont les substantifs les plus représentés que nous avons choisi d'extraire. Etant donné le peu de corpus de comparaison disponibles, le recours à une analyse des spécificités a dû être écarté – il nous semble en effet peu pertinent de nous fonder sur un corpus comme *Le Monde* par exemple pour déterminer les concepts spécifiques de la linguistique².

Si nous avons par ailleurs mis au jour, au sein d'une étude comparative des discours linguistique, philosophique et critique (Loiseau, Poudat et Ablali, 2006) certains items spécifiques à la linguistique, ils renvoient la discipline à ses observables (*verbe, phrase, énoncé, mot*, etc.), en excluant la *langue* ou le *sens*, que s'approprient également la philosophie et la critique ; il nous semblerait peu pertinent d'exclure certains objets parce que d'autres disciplines les empruntent.

Les variations flexionnelles (singulier/pluriel) ont été prises en compte, dans la mesure où le trait « nombre » n'indique pas seulement la pluralité : *langue* et *langues* sont ainsi deux concepts linguistiques distincts.

² Démarche pourtant adoptée par L'homme (2004) dans une approche terminologique intéressée par le domaine informatique.

Remarquons d'emblée qu'on relève trois fois plus de candidats concepts au singulier qu'au pluriel (75,77% sg. vs. 24.23% pl., avec un vocabulaire de 9680 formes au singulier, et 3811 formes au pluriel) et que certains substantifs semblent résolument singulier ou pluriel, tandis que d'autres paraissent plus indéterminés en nombre. Les hapax représentent respectivement 40,53 et 40.62% des formes relevées au singulier et au pluriel ; on observe d'ailleurs des paliers de fréquence comparables des deux types de substantifs :



Graphique : Paliers de répartition des substantifs au singulier et au pluriel (en pourcentages)

Si l'on ne relève que deux substantifs au pluriel de fréquence absolue supérieure à 1000 (*mots* et *verbes*), on en observe 22 au singulier – soulignons que les connecteurs comme *en ce sens* ou *de (en) fait* sont étiquetés sous d'autres catégories, et n'entrent pas en compte dans les relevés :

Rang	Substantif	Fréquence absolue	Textes ³
1	sens	2136	200
2	langue	2037	176
3	forme	1840	208
4	cas	1693	214
5	discours	1687	155
6	relation	1654	183
7	type	1650	205
8	verbe	1632	142
9	contexte	1568	157
10	objet	1500	191
11	sujet	1351	165
12	texte	1313	121
13	point	1297	210
14	analyse	1263	185
15	phrase	1237	143
16	rapport	1223	196
17	énoncé	1206	149
18	langage	1184	111
19	interprétation	1068	162
20	fait	1057	194
21	fonction	1031	187
22	français	1020	146

Graphique : Substantifs au singulier de fréquence absolue > 1000

On distingue globalement deux types de substantifs : les candidats concepts linguistique (en gris), et les substantifs relevant visiblement de la méthodologie scientifique à l'œuvre dans les articles. On observe ainsi un intérêt prononcé de la discipline pour le *sens*, qui détrône la *langue*, objet pourtant intuitivement premier de la linguistique. Le *discours* est également très honorablement représenté, tandis qu'on observe un intérêt particulier pour le *verbe*, dont la forme fléchie plurielle détient le second rang des substantifs au pluriel (1225 occ. pour 117 textes).

Les résultats obtenus confirment par ailleurs le centrage de la linguistique française sur la langue française – rappelons que peu de langues sont représentées dans le corpus (essentiellement l'anglais, l'espagnol, l'allemand, auxquelles on peut éventuellement rajouter l'ancien français).

Si le *texte* l'emporte sur la *phrase* en termes de fréquences absolues (1313 vs. 1237), on observe qu'il apparaît pourtant dans un nombre plus restreint d'articles (121 vs. 143), et que c'est finalement *l'énoncé* qui domine selon ce dernier critère (149 textes) : on voit là l'intérêt de prendre la fréquence et le nombre de textes d'apparition de l'occurrence.

Nous avons donc ordonné l'ensemble des substantifs au singulier et au pluriel en prenant en compte les deux paramètres, ce qui nous donne le classement suivant – on notera que le

³ Indique le nombre de textes dans lesquels le substantif a été relevé.

nombre de concepts a sensiblement diminué et que l'énoncé devance la phrase et le texte (26^e rang dans le nouveau classement) :

Rang	Substantif	Fréquence absolue	Textes
1	sens	2136	200
2	forme	1840	208
3	cas	1693	214
4	langue	2037	176
5	type	1650	205
6	relation	1654	183
7	objet	1500	191
8	point	1297	210
9	discours	1687	155
10	contexte	1568	157
11	rapport	1223	196
12	analyse	1263	185
13	verbe	1632	142
14	sujet	1351	165
15	fait	1057	194
16	fonction	1031	187
17	question	986	191
18	énoncé	1206	149
19	phrase	1237	143
20	partie	917	190

Graphique : Substantifs au singulier ordonnés par fréquence et par textes

Les substantifs restants relèvent de la logique (*relation, rapport*), de la typologie (*cas, type*) ou sont trop ambigus pour renvoyer directement à un objet linguistique (*sujet, objet, fonction, point*).

Voici le tableau rassemblant les 20 premiers substantifs au pluriel relevés dans le corpus :

Rang	Substantif	Fréquence absolue	Textes
1	mots	1293	143
2	exemples	991	173
3	formes	980	152
4	verbes	1225	117
5	cas	790	174
6	relations	956	139
7	types	717	169
8	termes	715	165
9	langues	960	118
10	énoncés	790	134
11	éléments	657	153
12	propriétés	787	121
13	unités	669	110

14	phrases	651	102
15	structures	478	113
16	textes	603	89
17	constructions	527	100
18	conditions	453	116
19	données	446	115
20	emplois	461	106

Graphique : Substantifs au pluriel ordonnés par fréquence et par textes

Bien que la plupart des candidats relevés au pluriel aient déjà été observés au singulier (e.g. *verbe(s)*, *langue(s)*, *énoncé(s)*), certains substantifs semblent résolument employés au singulier : ainsi, « sens » est le substantif singulier le plus relevé dans le corpus (2186 occ. au singulier) et il est notable qu'il soit globalement peu employé au pluriel (179 occ./83^e rang). Il en va de même pour *discours* (9^{ème} substantif), qui est relevé 1732 fois au singulier et seulement 146 fois au pluriel (118^e rang), ou encore pour *langage* (1208 occ. au singulier vs. 30 au pluriel).

S'il est intuitif que *le discours* et *les discours*, ou *le langage* et *les langages* ne renvoient pas aux mêmes concepts linguistiques, de tels écarts demeurent surprenants.

On ne note pas de différences aussi importantes à l'inverse, mais plusieurs substantifs qui relèvent davantage de la méthodologie que d'une thématique strictement linguistique, sont essentiellement pluriels : *données*, *conditions*, *caractéristiques*, *phénomènes*, *traits*, *critères*, *résultats* ou *contraintes*.

Soulignons que ce bref panorama des candidats concepts linguistiques n'aura finalement porté que sur les *hautes fréquences*. Malgré leur intérêt descriptif, les hapax et les éléments moins – ou plus inégalement – représentés ont été globalement écartés des analyses qui suivent, dans la mesure où ils se prêtent difficilement à l'analyse statistique.

6.2. Corrélations morphosyntaxiques

Nous avons ensuite retenu une sélection de substantifs parmi les hautes fréquences relevées, auxquelles ont été par hypothèse adjoints *sémantique*, *énonciation* et *cotexte* :

SENS : SENS :Nsg, SENS :Npl
SEMANTIQUE : SEMA :Nsg, Npl, SEMA :Asg, Apl
LANGUE : LGUE :Nsg, Npl
DISCOURS : DISC :Nsg, Npl, Asg, Apl (discursif (ve) (s))
PAROLE : PAR :Nsg
LANGAGE : LANG Nsg, Npl, Asg, Apl (langagier, langagière (s))
TEXTE : TXT : Nsg, Npl, Asg, Apl (textuel(le)(s))
CORPUS : CORP Nsg, Npl
INTERPRETATION : INTER Nsg, Npl, Asg, Apl (interprétatif, interprétative)
CONTEXTE : CONT Nsg, Npl, Asg, Apl (contextuel(le)(s))
COTEXTE : COT Nsg, Npl
ENONCIATION : ENONC Nsg, Npl, Asg, Apl (énonciatif(ve)(s))

Dans la mesure où elles sont fondées sur le même morphème, les formes adjectivales des entrées ont également été prises en compte.

Les corrélations des entrées entre elles et avec le jeu de descripteurs morphosyntaxiques ont été observées et c'est sur les textes entiers (avec exemples et citations) qu'ont été effectuées les analyses. L'entreprise peut paraître discutable, mais il s'avère que si l'on extrait les exemples du corpus, il n'est plus possible de voir si les formes sont employées dans des textes contenant des exemples, ceux-ci étant particulièrement identifiables (v. chapitre 5).

Etant donné les différences d'échelle des deux types de données⁴, les substantifs sont globalement corrélés entre eux et peu de corrélations négatives significatives (seuil > -0.2) ont pu être relevées. A l'inverse, plusieurs corrélations positives très élevées (> +0.7) ont pu être observées.

De manière générale, les corrélats des candidats observés diffèrent au singulier et au pluriel, ce qui suppose qu'ils renvoient à des objets distincts. La différence est particulièrement visible si l'on s'intéresse à *sens* : si les deux formes sont corrélés entre elles (+0.33), leurs corrélats diffèrent d'abord en nombre, eu égard à la représentation dix fois plus élevée de *sens* au singulier (14 corrélations significatives au singulier vs. 6 au pluriel).

Sens au singulier est fortement corrélé à *contexte* et à *interprétation* (respectivement +0.47 et +0.41), et ses corrélats sont d'ailleurs essentiellement des candidats concepts au singulier : l'entrée est ainsi corrélée au *contexte* et non aux *contextes*, au *texte* et non aux *textes*. En outre, *sens* au singulier est négativement corrélé aux marques de formalisation (symboles, abréviations linguistiques, etc.), aux numéraux et aux parenthèses, caractéristiques des textes plus appliqués : le concept serait ainsi plus représenté dans les textes à dominante théorique.

Ces données conduisent à penser que la forme *sens* au singulier renvoie bien à une lexicalisation privilégiée de concept ; au contraire, la forme plurielle de *sens* est corrélée aux marqueurs caractéristiques des textes plus appliqués – et plus exemplifiés (pronom personnel *tu*, symboles linguistiques ?, *, ! et # et connecteurs d'exemplification). En ce sens, *sens* au pluriel serait une forme peu discutée, voire instrumentale.

On observe un phénomène similaire pour les deux formes singulier et pluriel de *corpus* – qui partagent d'ailleurs les mêmes corrélats : si on les observe aux côtés des *textes* et des *discours*, les deux formes sont corrélées à différentes caractéristiques des textes plus exemplifiés (interjections, connecteurs d'exemplification, etc.). *Corpus* serait ainsi un objet instrumental, au même titre que *sens* au pluriel.

De manière non surprenante, la *langue* et les *langues* ne renvoient pas aux mêmes objets : corrélée à la *parole* et au *langage* (substantifs et adjectifs), la *langue* semble discutée dans des textes à dominante historique, dans la mesure où elle est corrélée au temps de l'imparfait 0.23 (et du plus-que-parfait), eux-mêmes fortement corrélés au passé simple et aux dates 0.2. A fortiori, elle s'oppose à de nombreuses caractéristiques des textes scientifiques (symboles, numéraux, marqueurs de structuration des textes, impératif et présent). Il en va fort différemment des *langues*, corrélées aux *langages* et au *sémantique* (substantifs et adjectifs), pour lesquelles on ne repère pas cette dimension historique.

Les corrélats morphosyntaxiques semblent ainsi représenter un critère efficace pour discriminer les candidats et les dimensions textuelles, voire même les pôles génériques (v. chapitre 3) auxquels ils sont associés (textes historiques/exemplifiés/formels, etc.) ; notons toutefois que les résultats obtenus ne sont pas tous aussi probants, les candidats observés

⁴ Fréquences absolues des substantifs vs. fréquences relatives des variables morphosyntaxiques.

ayant des fréquences inégales. *Langue(s)* et *sens* sont ainsi particulièrement représentés et stabilisés dans le corpus.

On soulignera enfin l'intérêt de ne pas considérer le *lemme*, mais les *formes fléchies*, et plus précisément le trait *nombre* des objets observés, qui s'avère discriminer certains concepts entre eux.

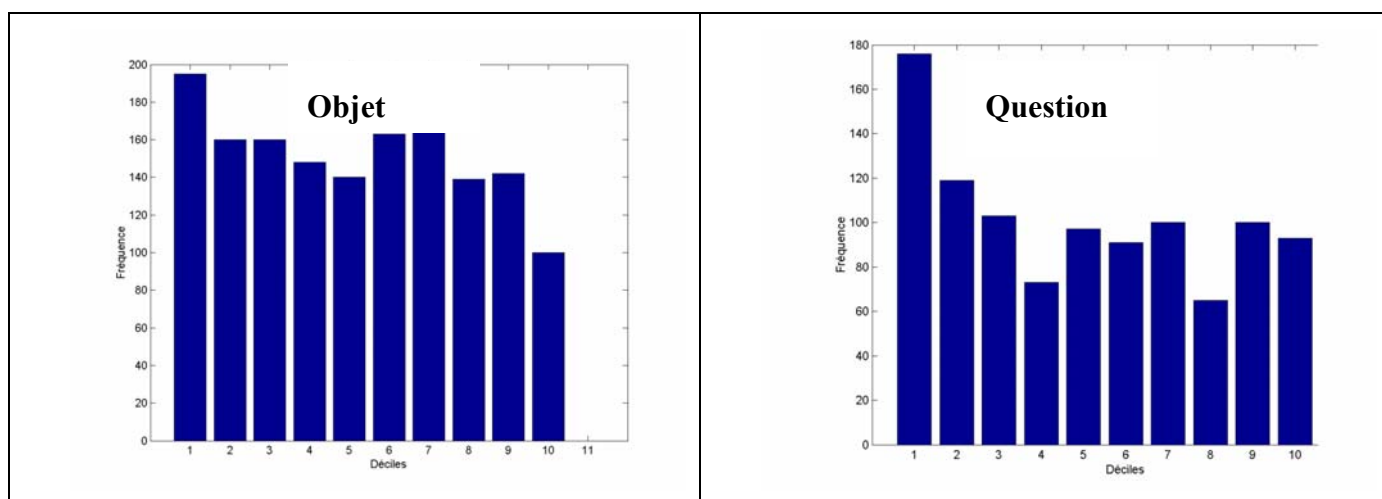
6.3. Tactique

De la même manière que nous avons apprécié la distribution des variables morphosyntaxiques dans la séquentialité des textes du corpus, les concepts ont été observés au sein de leurs déciles textuels, au moyen du logiciel CR développé par S. Loiseau⁵ - cette représentation nous a finalement semblé plus adaptée que l'observation des candidats au sein de leurs sections textuelles, globalement inégales (v. chapitre 4).

Les profils obtenus sont particulièrement discriminants : certains concepts sont plus concentrés en début, en corps ou en fin d'article, tandis qu'on observe des profils caractéristiques qui pourraient bien signifier une problématisation de l'objet.

6.3.1. Concepts de début / fin d'article

Certaines entrées, comme *objet* ou *question* sont plus concentrées en début d'article :

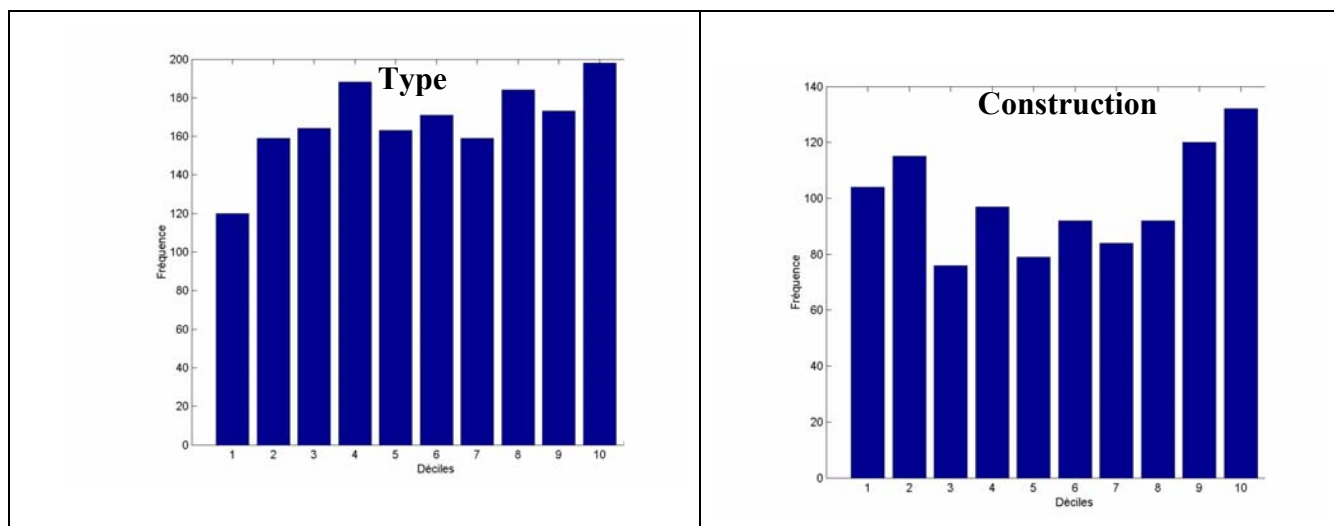


Graphique : Configurations tactiques de OBJET et QUESTION (partitionnements en déciles)

On observe ainsi une décroissance des deux concepts de l'introduction à la fin du texte. Dans la mesure où ils sont plus concentrés en début d'article, on peut penser qu'ils participent à la problématisation de la recherche présentée ; *objet* et *question* seraient alors des concepts instrumentaux plutôt que discutés.

On relève au contraire un emploi plus important de *type* et de *construction* en fin d'article :

⁵ <http://panini.u-paris10.fr/~sloiseau/CR/>



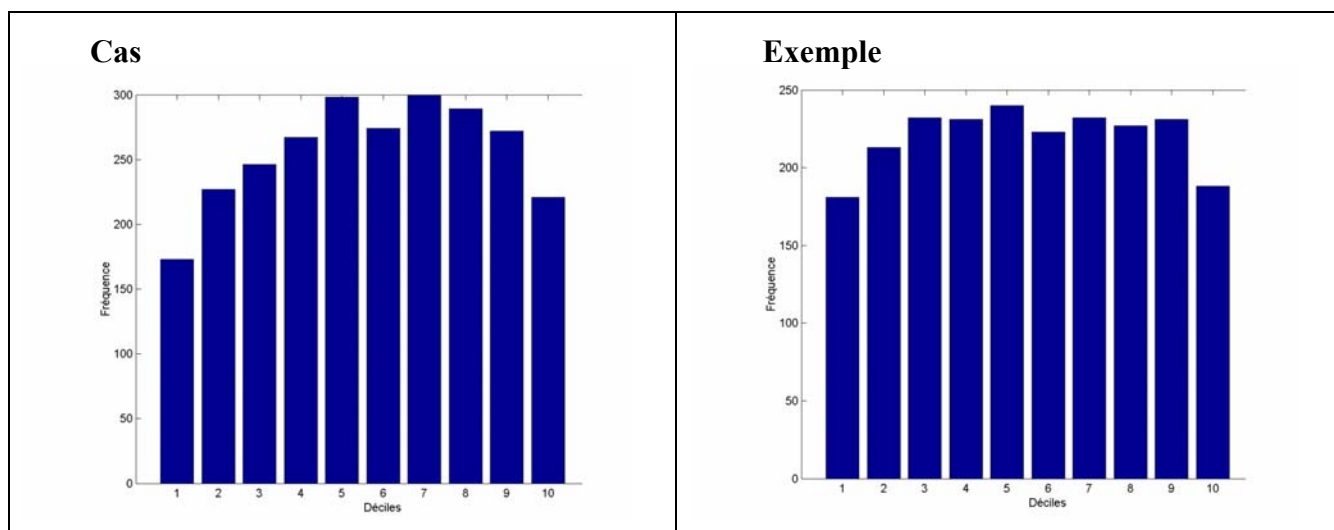
Graphique : Configurations tactiques de TYPE et CONSTRUCTION (partitionnements en déciles)

Une telle forme croissante des concepts pourrait renvoyer aux objectifs généraux de la démarche scientifique linguistique – ici classificatoires et typologiques.

Il serait ainsi intéressant de déterminer de manière plus précise et plus exhaustive les objets de début et de fin d'article, qui semblent plus méthodologiques que véritablement discutés. On pourrait ainsi les contraster d'une discipline scientifique à l'autre, afin de comparer les démarches et les présupposés méthodologiques adoptés.

6.3.2. Concepts de corps d'articles

Si certains objets semblent d'abord liés à la problématisation ou aux résultats, on observe des concepts de corps d'article, comme *cas* ou *exemple* :

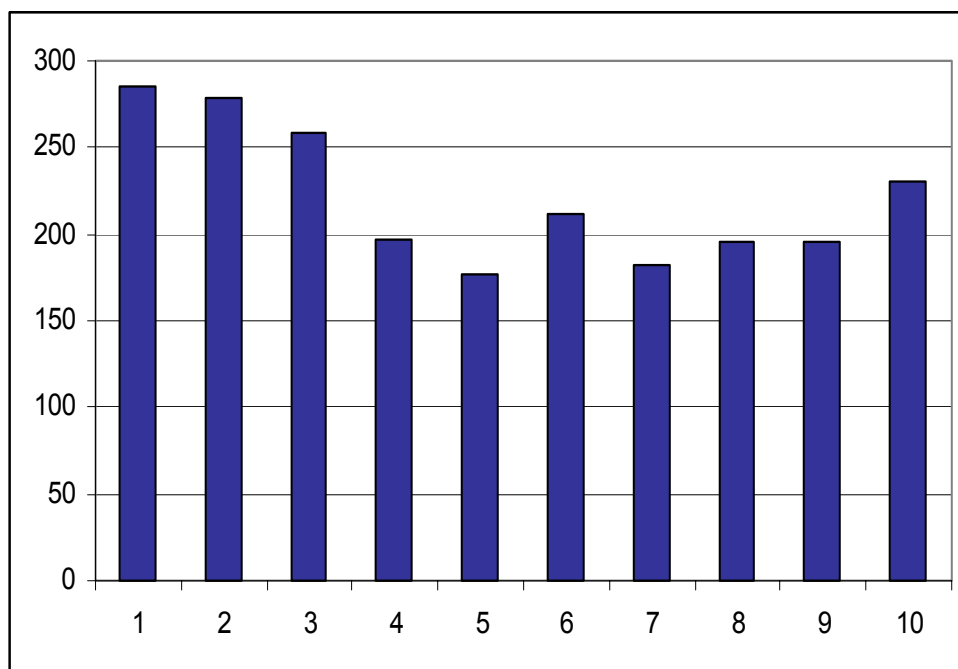


Graphique : Configurations tactiques de CAS et EXEMPLE

Ces objets sont nettement méthodologiques : leur forme est graduelle avec un double mouvement croissance / décroissance et un maximum obtenu en milieu d'article : non problématisés et non discutés en fin d'article, ces éléments sont essentiellement corrélés au développement et aux analyses menées.

6.3.3. Concepts problématisés

La configuration tactique qui a particulièrement retenu notre attention a une forme précisément inverse de celle des concepts de corps d'article :



Graphique : Configuration tactique de SENS

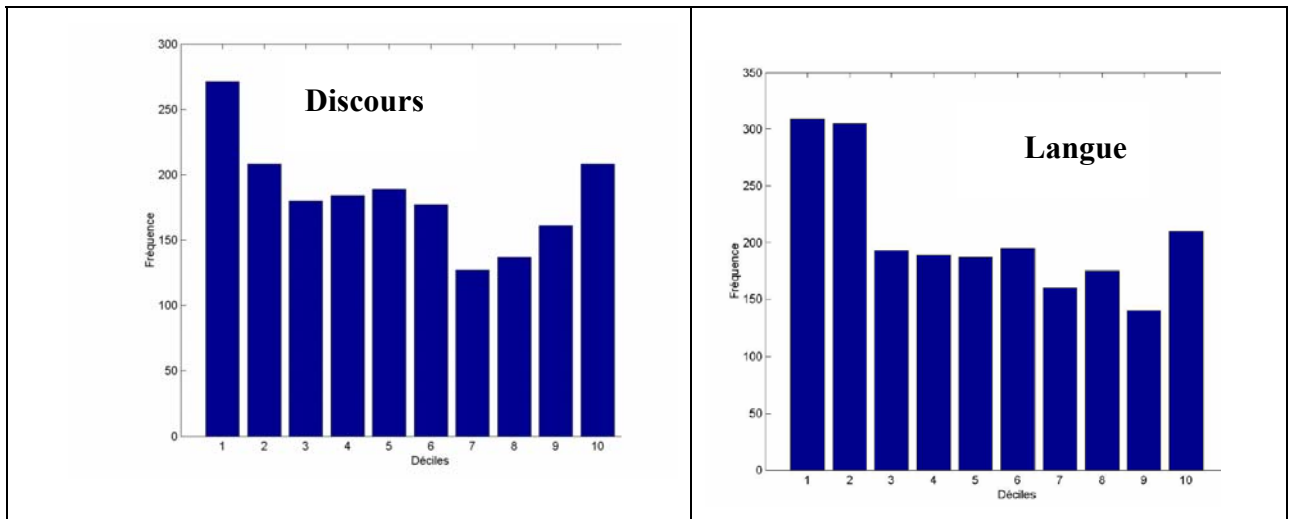
Sens est par exemple plus concentré en début et en fin d'article : on note une décroissance régulière du concept dans les trois premiers déciles, qui évoque un passage du général au spécifique (à partir du décile 4). La tendance s'inverse au-delà du décile 6, jusqu'au dixième décile – qui correspond globalement à la conclusion de l'article, où le concept revient brusquement, de manière vraisemblablement rhétorique.

Cette forme tactique incurvée semble spécifique aux concepts débattus, qui seraient ainsi davantage représentés en début et en fin d'article qu'en son corps :

Généralisation/problématisation (maximum atteint) → spécification → retour conclusif

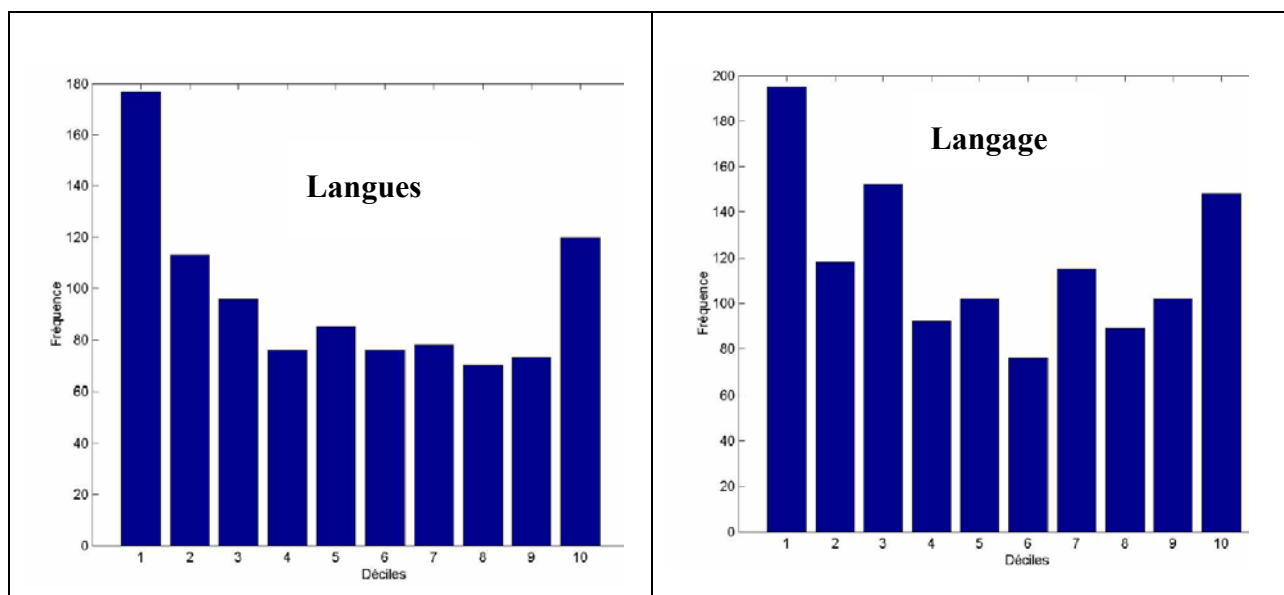
Reconsidérons la liste des 20 substantifs de hautes fréquences (et les plus également répartis dans le corpus) mise au jour en début de chapitre : *sens, forme, cas, langue, type, relation, objet, point, discours, contexte, rapport, analyse, verbe, sujet, fait, fonction, question, énoncé, phrase* et *partie*.

Outre *sens*, seuls deux candidats satisfont le critère tactique qui nous intéresse : *discours* et *langue* :



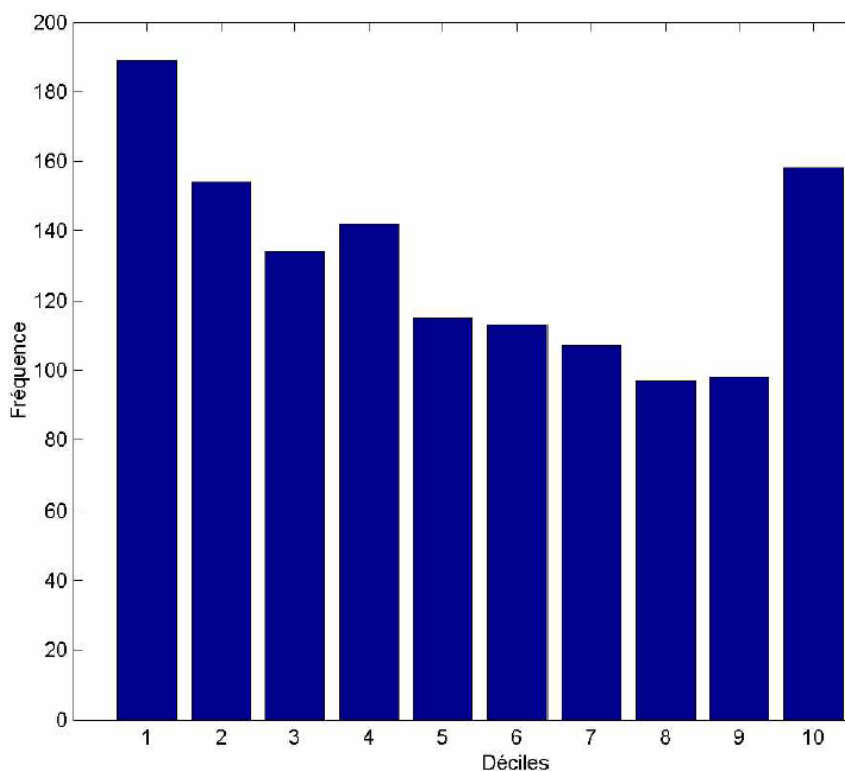
Graphique : Configurations tactiques de DISCOURS et LANGUE

Malgré des différences de répartition des deux concepts dans le développement de l'article, on observe bien un retour sur le concept en fin d'article et un pic de représentation maximal en début d'article : *discours* et *langue* seraient bien des concepts de fond disciplinaire, et il en va d'ailleurs de même pour *langues* et *langage* :



Graphique : Configurations tactiques de LANGUES et LANGAGE

Si cette organisation tactique combinée au critère de fréquence nous semble ainsi permettre de discriminer les concepts discutés des autres, mentionnons une irrégularité, figurée par le substantif de haute fréquence *analyse*⁶ :



Graphique : Configuration tactique de ANALYSE

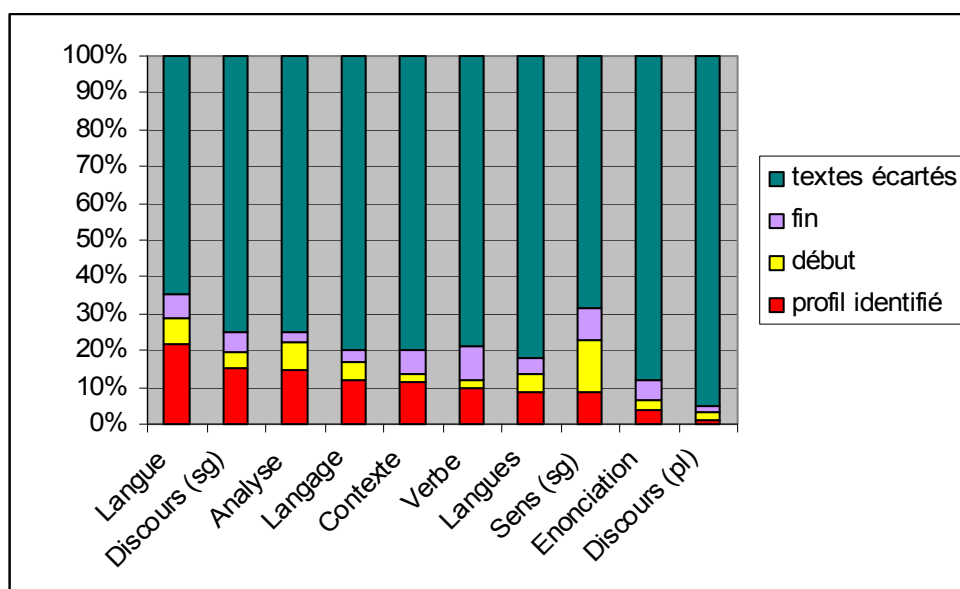
⁶ Qui est néanmoins aussi un concept hjemslevien.

Les résultats obtenus figurant davantage des formes textuelles génériques que des formes textuelles individuelles – les décomptes étant effectués au niveau du corpus -, nous nous sommes dans un deuxième temps attachée à vérifier l’existence de ces formes et leur statut d’objet discuté dans les textes du corpus.

Les configurations tactiques de chaque texte ont été observées manuellement : les résultats obtenus n’ont qu’une valeur d’approximation, l’identification des formes étant souvent délicate – il serait à terme pertinent de développer un module d’identification automatique des configurations tactiques.

Aux six concepts mis au jour nous avons adjoint à titre illustratif quatre concepts de configuration tactique globale distincte : la forme plurielle de *discours*, *énonciation*, qui s’avère un concept de fin plutôt que de début d’article, *contexte* et *verbe*, pour lesquels nous n’avons détecté aucune forme tactique particulière.

Les résultats obtenus montrent que *langue* est le concept le plus débattu : 20% des textes qui contiennent l’entrée ont la configuration tactique qui nous intéresse, contrairement à *sens*, qui s’avère comparativement plus employé en début de texte :



Graphique : Configurations tactiques des formes les plus représentées du corpus

A fortiori, les articles ayant la configuration tactique précédemment observée semblent effectivement discuter la notion : le critère appliqué à l’objet *langage* permet par exemple d’identifier tous les textes contenant le concept dans leur titre et, de manière plus intéressante, les textes qui discutent la notion sans qu’elle soit nécessairement annoncée. Par exemple, on relève la forme tactique pour le concept de *langue*, qui est en effet discutée dans l’article de D. Leeman (*Dans un juron, il sauta sur ses pistolets. Aspects de la polysémie de la préposition dans*, texte 032), ce qui n’aurait pas nécessairement été mis au jour sans ce critère.

Ce paramètre tactique, de mise en œuvre plus aisée que d’autres critères, semble ainsi particulièrement discriminant, et pourrait intéresser certaines applications de recherche d’information, en facilitant le repérage et la localisation des thèmes textuels.

6.4. Co-occurents lexicaux

L'examen des co-occurents lexicaux est particulièrement crucial dans le processus de qualification thématique des candidats concepts : si leur nombre implique une plus ou moins grande stabilisation de la forme et si les écarts observés mesurent le degré de corrélation contextuelle entre les mots, leurs éventuelles intersections sémiques – qui indiquent la présence d'isotopies – les fait accéder au statut de *corrélats sémantiques* (Rastier).

Dans la mesure où il n'est pas envisageable d'analyser les co-occurents de l'ensemble des candidats concepts du corpus, nous avons d'abord choisi de nous concentrer sur les objets *sens* et *langue*, choix motivé par leurs hautes fréquences et leur qualité d'objets / objectifs de descriptions linguistiques privilégiés.

Notons d'abord qu'en dépit de leurs fréquences distinctes (2471 occ. de *sens* vs. 1798 occ. de *langue*), les deux formes ont quasi le même nombre de co-occurents : 209 pour *sens* vs. 207 pour *langue*. On relève des écarts de corrélation plus importants pour *langue* que pour *sens* : *langue* draine ainsi plus de lieutenants stabilisés que *sens*.

Précisons que pour observer l'objet *langue*, nous avons dû écarter le numéro de *Langue française* sur la *langue des signes* dont nous disposons, *langue* apparaissant d'abord corrélée aux spécificités de ce numéro : *signes, sourds, LSF, entendants*, etc.

Ecart	Corpus	Extrait	Mot
160.80	1798	1832	LANGUE
24.33	80	60	PARLÉE
22.41	39712	3801	LA
18.39	136	63	FRANÇAISE
16.98	1329	245	LINGUISTIQUE
16.14	50	32	SPÉCIALITÉ
15.80	33	25	USUELLE
15.52	34	25	MATERNELLE
15.09	62212	5144	DE
14.95	99	44	SAUSSURE
14.54	30	22	ÉTRANGÈRE
14.30	736	147	SYSTÈME
13.50	913	164	LANGUES
12.98	21503	1929	DES
12.65	636	123	LINGUISTIQUE
12.41	27	18	VIVANTE
12.11	123	42	NATURELLE
11.12	19087	1674	UNE
10.61	25	15	TRÉSOR
10.16	37	18	SPÉCIALISÉE
9.95	24123	2016	ET
9.84	237	54	LEXIQUE

Tableau : Premiers co-occurents de LANGUE – ASL

On relève ainsi pour *langue* les figements :

langue parlée, française, maternelle, naturelle, étrangère, usuelle, de spécialité, courante, Trésor de la langue française, langue vivante, anglaise, cible, spécialisée, allemande, donnée, poétique, orale, écrite, standard, système de la langue

eux-mêmes corrélés à des co-occurents et des contextes spécifiques :

- contexte historique (dialogue/russie/poésie) pour *langue parlée* ;
- contexte dictionnaire + politique linguistique pour *langue française* (TLF, Larousse, dictionnaire(s), Robert, etc. / *respectent, ministère, aménagement, historique, etc.*) ; on notera que l'on n'observe pas une telle stabilité pour les *langue(s) anglaise et allemande*, la première étant d'abord corrélée à l'Europe, la seconde à la guerre et aux allemands (!).
- contexte sociolinguistique pour *langue maternelle* (*alsacien, parisienne, âgés, supérieures, village, enregistrés, adultes, jeunes, région, etc.*).
- contexte informatisé pour *langue naturelle* (*interfaces, ordinateur, informatique, étiqueté, baser, langages, modèles, mécanismes, etc.*)
- contexte didactique pour *langue étrangère* (*restructuration, apprenants, acquisition, écriture, etc.*)
- emploi quasi indistinct de *langue courante/usuelle* (mêmes co-occurents)
- distinction *langue de spécialité* (contexte de traduction) et *langue spécialisée* (contexte informatique)
- si la *langue cible* est bien sûr associée à la traduction, on relève toutefois différentes catégories de co-occurents selon le type de pratique et de conception dans lequel est abordé le processus (didactique, traductologie, histoire...)
- *langue écrite* et *langue orale* ne sont visiblement pas des figements conceptuels, ils ne valent que par leur opposition.

On voit là toute la difficulté que pose l'observation des formes de haute fréquence, qui drainent de nombreux figements qui s'autonomisent de la notion première.

On observe le même phénomène pour *sens*, qui n'a qu'un rôle de formant dans de nombreux figements : *sens littéral / sens figuré, sens strict / sens large, sens commun...* auxquels on peut adjoindre *immanence du sens, sens distributionnel, donation du sens, [unité] porteuse de sens, compositionnalité du sens...* si l'on restreint le contexte pris en compte du paragraphe aux 50 caractères avoisinant la notion.

Ecart	Corpus	Extrait	Mot
166.06	2471	2529	SENS
17.94	48	39	FIGURÉ
15.62	1599	313	CONTEXTE
15.37	14256	1743	DU
15.09	29259	3245	LE
14.48	6702	910	AU
13.56	1045	213	MOT
13.43	25	21	LITTÉRAL
13.26	64296	6482	DE
13.17	1656	293	SÉMANTIQUE
12.27	771	162	TERME
12.08	428	107	SIGNIFICATION
11.84	7872	973	OU
10.77	20555	2204	UN
10.40	1354	224	TEXTE
10.19	218	61	COMMUN
9.81	1395	223	MOTS
9.70	8666	1000	CE

9.63	84	32	ING
9.28	151	45	SIGNIFIÉ
9.27	167	48	LARGE
9.17	31	17	STRICT

Tableau : Premiers co-occurents de sens – corpus ASL 224 – contexte paragraphe

Notons que les deux objets sont corrélés à *Saussure*, et que la *langue* est corrélée à la dichotomie saussurienne *langue / parole* (*parole* et *Saussure*) tandis que *sens* s’inscrit dans la problématique du *signe* (*signifié, signe, signifiant, Saussure*).

On observe enfin que *sens* est associé aux domaines lexical (*mot(s), lexical, terme...*) et textuel (*texte*), et qu’il porte un sème [+ nouveau][+ changement] : *nouveau sens, changement de sens*.

6.5. Incidence du numéro thématique

6.5.1. SENS Revue Contexte(s)/ASL hors revue Contexte(s)

Ecart	Corpus	Extrait	Mot
34.66	289	291	SENS
6.06	9	9	CLASSIQUE
5.60	10	9	ÉTABLISSEMENT
5.50	1524	385	LE
5.50	23	15	CONTENU
5.35	7	7	LITTÉRAL
5.33	680	189	CONTEXTE
5.19	305	96	AU
4.80	10	8	VIRTUEL
4.77	24	14	COMMUN
4.77	19	12	HORS
4.64	86	34	NOTION
4.51	28	15	MESSAGE
4.46	120	43	MOT
4.43	40	19	APPROCHE
4.39	9	7	ASSEMBLAGE
4.10	12	8	DÉTERMINATION
4.10	12	8	COMPRENDRE
4.08	49	21	SUJET
4.04	4	4	CENTRALE
3.95	940	233	UN
3.94	8	6	ENJEU

Tableau : Premiers co-occurents de sens – corpus revue contexte(s)

La présence de *contexte* dans les premiers co-occurents de *sens* dans le corpus global s’avère en grande partie liée au numéro *Contexte(s)*. *Sens* était d’ailleurs plus fortement co-occurent à *contexte* au niveau du paragraphe (deuxième co-occurent relevé, derrière *figuré*). Si l’on ne prend en compte que la revue *Contexte(s)*, *contexte* est le cinquième co-occurent de *sens*, tandis qu’il n’est que le 63^e lorsqu’on considère le reste du corpus d’articles.

Les co-occurents de *sens* diffèrent fortement aux niveaux de la revue et du reste du corpus : on ne retrouve pas la plupart des items obtenus d’un corpus à l’autre. Le concept de

sens dans la revue observée ne semble pas débattu en tant que tel, mais relativement à la notion de contexte ; en d'autres termes, on ne s'intéresse au *sens* que par rapport au *contexte*, d'où le figement *sens littéral* (en contexte zéro) et la récurrence de la notion de *contenu* (topos : le sens n'a pas de contenu sans prise en compte du contexte) :

Lakoff et Johnson (La métaphore dans la vie quotidienne 1981, p. 21) montrent, à travers la métaphore du conduit, que l'idée d'un **sens contenu** dans une forme linguistique est inhérente à notre perception de la langue : on parle de « faire passer, donner une idée, introduire une idée dans une phrase, d'une phrase vide de **sens**. » (Les contextes de contexte La notion de contexte dans les Page: 24 c (41ème occ.))

On a vu que les IC n'ont pas de **sens référentiel**, décontextualisé. Ils sont plus « flexibles » que les signes linguistiques dans la mesure où ils n'ont pas de **contenu** informatif stable. L'interprétation est fonction de ce qui se passe dans l'interaction, et non de la nature de l' IC en tant que tel. (Texte contexte contextualisation : A-t- on progressé Page: 84 b (74ème occ.))

[...] il importe de dégager les processus qui permettent l'établissement d'un **sens en soi**, **contenu** dans la langue, indépendamment du contexte. (Production et interprétation du sens : la notion de contexte Page: 279 c (179^{ème} occ.))

On observe également une co-occurrence de *sens* et de la préposition *hors*, qui renvoie encore une fois au contexte, sous des formes différentes (*hors contexte*, *hors circonstance*, *hors langue*, etc.).

6.5.2. LANGUE Revue Contexte(s)/ASL hors revue Contexte(s)

Ecart	Corpus	Extrait	Mot
34.65	117	120	LANGUE
8.81	8	8	SAUSSURE
8.24	7	7	ALLEMANDE
7.05	9	7	FRANÇAISE
6.04	1750	237	LA
6.04	79	23	SYSTÈME
5.93	18	9	FAITS
5.86	289	56	SENS
5.47	32	12	SYSTÈMES
5.40	3	3	FICELLE
5.40	3	3	DONNÉS
5.27	25	10	GUERRE
4.82	6	4	ALPHABÉTIQUE
4.76	9	5	CODE
4.56	13	6	MATERNELLE
4.51	4	3	ROBERT
4.51	4	3	INSTITUTION
4.51	4	3	GLLF
4.51	4	3	CENTRALE
4.35	7	4	ENTOURE
4.19	23	8	CONTENU
4.12	11	5	ÉCRITE

Tableau : Premiers co-occurents de LANGUE – ASL hors numéro Contexte(s)

Notons d'emblée la co-occurrence *sens / langue* ; dans les contextes observés, c'est en effet aux côtés de la *langue* que le *sens* est défini :

D'une part, l'idée selon laquelle le contexte modifie, éclaire ou sature les unités suppose celle d'un **sens** lexical de départ, c'est-à-dire fixé au niveau de la **langue** et façonnable (Les contextes de contexte)

Lakoff et Johnson (*La métaphore dans la vie quotidienne* 1981, p. 21) montrent, à travers la métaphore du conduit, que l'idée d'un **sens** contenu dans une forme linguistique est inhérente à notre perception de la **langue** (ibid.)

CONTEXTE. 1. Ensemble du texte qui entoure un élément de la **langue** (mot, phrase, fragment d'énoncé) et dont dépend son **sens**, sa valeur. (Texte et contexte)

En effet, si nous tentons d'interpréter Saussure selon le critère de la cohérence théorique, nous remarquerons d'abord qu'il définit la **langue** comme « un système de signes où il n'y a d'essentiel que l'union du **sens** et de l'image acoustique ... » et, que ce rapport de signification repose sur une convention. (Contexte et circonstances dans la sémiotique juridique)

Contrairement à *sens*, qui semble être un concept finalement caméléon, dont les co-occurents – et les acceptions – varient selon l'objet observé, la *langue* est un concept relativement plus stabilisé, qui semble invariablement porter les traits [Saussure], [système] [française] et [code]. On observera toutefois que l'objet *contexte* ne semble pas spécifiquement recourir à la dichotomie *langue / parole*, la *parole* étant absente des co-occurents de *langue* dans l'ensemble du numéro.

6.6. Incidence de la variable Auteur

On a ensuite cherché à observer la stabilité des concepts *sens* et *langue* d'un auteur à l'autre, à partir d'un corpus de 12 auteurs de linguistique (Authier, Barbéris, Bergounioux, Combettes, François, Kerbrat, Kleiber, Moirand, Neveu, Rabatel, Rastier et Siblôt).

6.6.1. SENS

6.6.1.1. Différentiel ASL/Auteurs

Malgré plusieurs différences, on relève de fortes régularités corrélatives (en gris dans le tableau ci-dessous), qui suggèrent un fond commun – rappelons que les corpus ASL et Auteurs sont globalement bien distincts, et ne partagent qu'une dizaine de textes :

Ecart	Corpus	Extrait	Mot
126.52	1551	1604	SENS
16.02	510	154	MOT
15.08	128	62	SIGNIFICATION
14.77	860	208	X
14.46	10653	1443	DU
13.15	755	177	SÉMANTIQUE
12.87	4797	714	AU
12.46	268	85	NOMS
12.43	26	21	DÉNOMINATIF
12.09	165	61	PRODUCTION
12.09	97	44	SIGNIFIÉ
11.84	394	106	TERME
11.58	14	14	LITTÉRAL
10.96	29	20	STRICT
10.75	47	26	STABLE
10.62	267	76	DÉNOMINATION
10.53	1825	304	DIRE
10.45	88	37	LEXICAL
9.84	18619	2152	LE

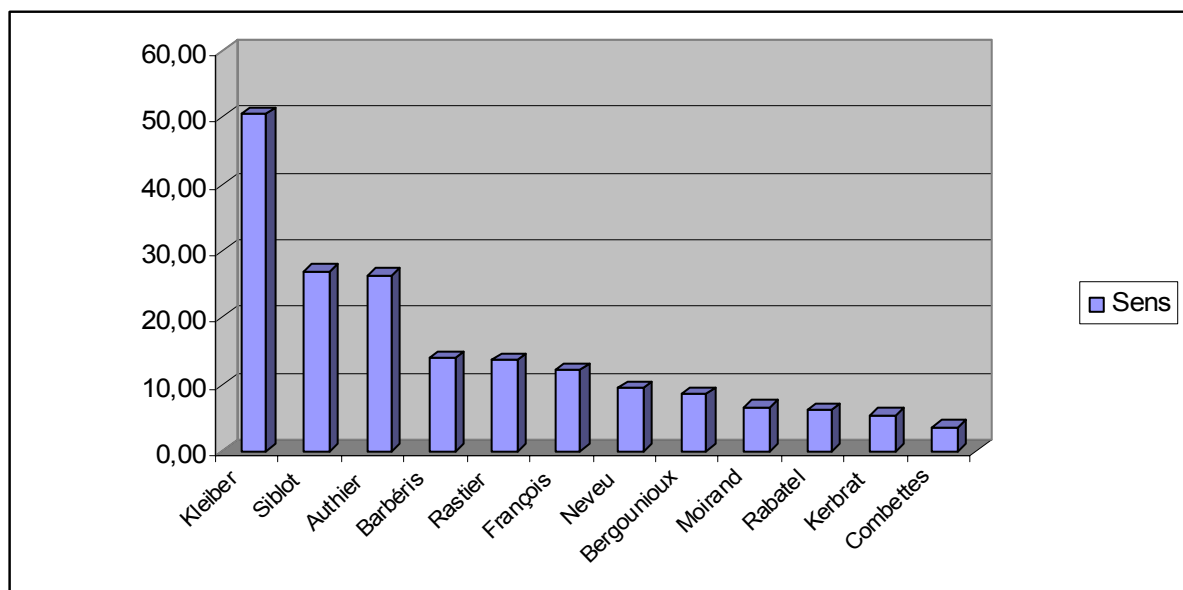
9.63	655	134	MOTS
9.60	239	66	PROPRES
8.69	22	14	GLISSEMENTS

Tableau : Premiers co-occurents de sens – ASL hors revue contexte(s)

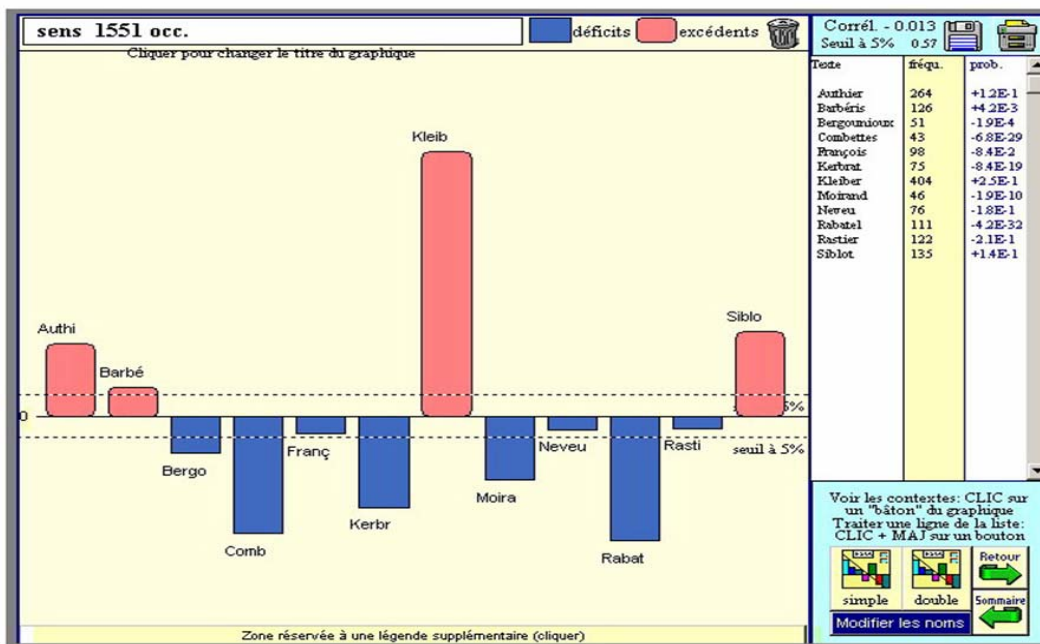
S'il s'inscrit dans le *sémantique*, le lexème *sens* semble régulièrement et prioritairement dédié à l'observation de l'objet *mot*, qui demeure le premier co-occurent partagé par les deux corpus (16.02 ASL – 1^{er} rang, 13.56 Auteurs – 2^e rang). On distingue deux contextes d'emplois principaux : les environnements lexicaux et terminologiques (*terme(s)*, *lexical(s)*, *lexicales*, *lexicaux*, *lexique*) et l'inscription de *sens* dans la problématique du signe (*signifié*, *signe(s)*, *signifiant*). L'objet *sens* est également corrélé, mais de manière moins significative, à la *langue*, au(x) *texte(s)* et au *contexte*, et aux questions de la *référence*, de la *polysémie* et de la *dénomination*.

On remarque en outre que *sens* s'inscrit dans un fond visiblement théorique (*conception*, *concept*, *conceptuel*, *idée*, *théorie(s)*, *notion*), ce qui corrobore les résultats obtenus dans les sections précédentes (2. et 3.).

6.6.1.2. Quantitatif



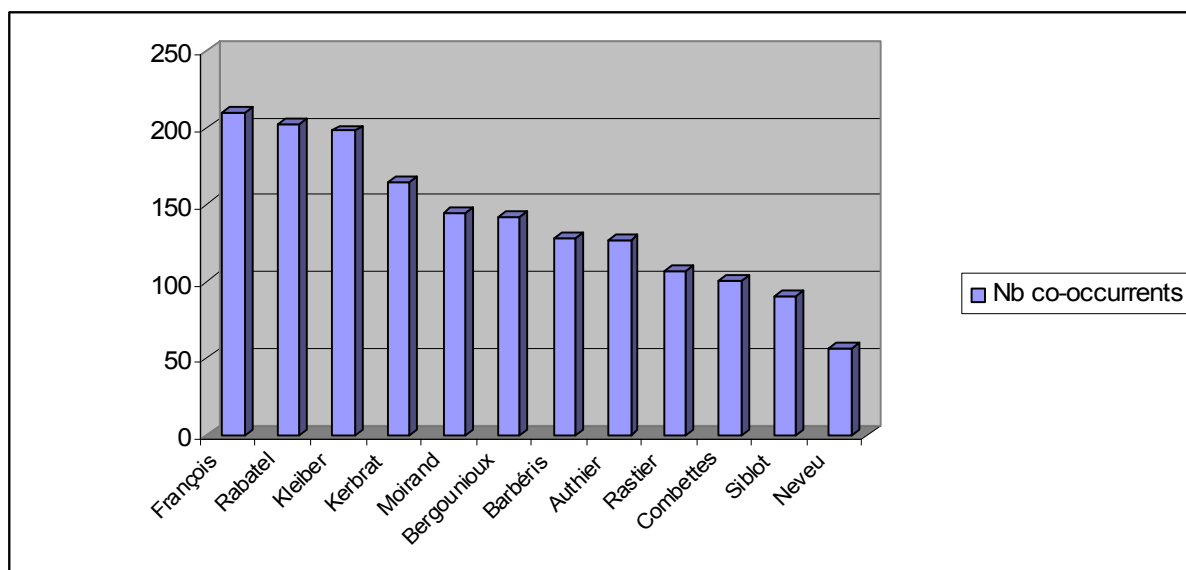
Graphique : Répartition de SENS (moyenne absolue par texte)



Graphique : Répartition de SENS (Hyperbase)

Si on relève des occurrences de *sens* dans l'ensemble des textes pris en compte, le lexème est inégalement employé d'un auteur à l'autre : Kleiber y recourt par exemple treize fois plus que Combettes. Le lexème paraît en outre inégalement distribué, et plus concentré chez certains auteurs : on relève plusieurs occurrences de *sens* au sein d'un même paragraphe chez Authier, Kleiber, Siblot et Moirand.

Bien que le nombre de co-occurents varie bien entendu avec celui du nombre de textes de départ (inégal d'un auteur à l'autre), le graphique qui suit fournit une approximation de la diversité des contextes d'apparition de *sens* ; ainsi, on dispose du même nombre de textes pour François et Neveu, positionnés pourtant aux deux extrémités du diagramme obtenu :



Graphique : Nombre de co-occurents de SENS

Objectif ou objet premier de la plupart des textes observés, le *sens* est diversement approprié et débattu d'un auteur à l'autre. On le trouve faiblement caractérisé chez les deux tiers des auteurs observés (Bergounioux, Combettes, Kerbrat, Kleiber, Moirand, Neveu et Siblot) : aucune corrélation supérieure à 10 n'est en effet observée – ce qui ne signifie pas que le lexème ne soit pas contextuellement régulé.

En revanche, on observe une stabilisation plus prononcée du lexème chez les quatre auteurs restants :

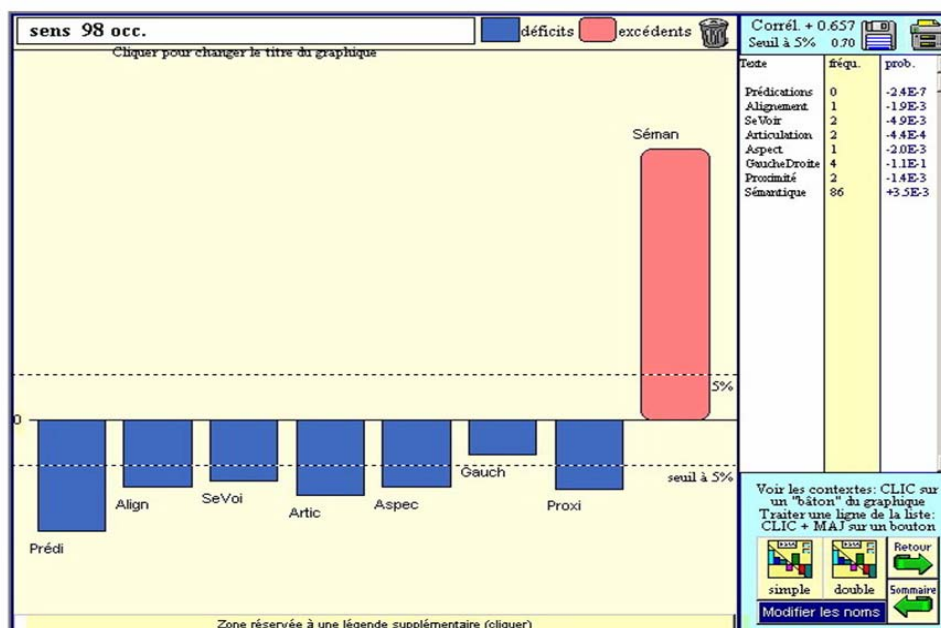
Authier	Barbérís	François	Rastier
X (15.23)	production (11.24)	contresens (14.23)	signification (15.59)
Q (14.52)		glissements (11.9)	
au (10.45)			

Tableau : Co-occurents de *sens* chez Authier, Barbérís, François et Rastier

- ✓ *sens* chez Authier apparaît régulièrement au sein de séquences reformulatives, ou explicatives de type *au sens de*, ou *au sens strict*. L'auteur s'intéresse en effet au sens d'un point de vue énonciatif, et cherche à saisir le sens de phénomènes tels que l'allusion d'un point de vue formel.
- ✓ Figement praxématique *production de sens* dans la perspective praxématique de Barbérís (présent dans l'ensemble des articles) ; *sens* est ainsi le formant d'un autre concept.
- ✓ Figement *glissements de sens* chez François et emploi co-occurent de *sens* et *contresens* (exclusivement relevés dans le texte *La sémantique*)
- ✓ Le *sens* est fortement stabilisé chez Rastier, et essentiellement distingué de la *signification*.

6.6.1.3. Analyse des co-occurents de *sens*

- **Authier** : les co-occurents de *sens* chez Authier semblent porter la trace d'un sème [+instable], lié à la perspective énonciative, d'où *risque*, (*référence*) *actuelle*, *ponctuellement*, *fixité (du signal)*, *équivoque*, *paraphrasable*, *polysémie*, *effets (de sens)*, etc.
- **Barbérís** : *sens* s'inscrit ici dans une conception praxématique et il articule les objets linguistiques (*préposition*, *complément*, *verbes*, *lexical*) au praxématique (*production*, *programmes*, *état*, *producteur*, *gestalt*, *praxémique* etc.).
- **Bergounioux** : *sens* s'inscrit dans la problématique du *signe* chez Bergounioux, qui le distingue du *signifié* (8.98, premier co-occurent observé) ; on observe ainsi un réseau conceptuel de corrélats articulés autour de cette question : *union* (signifiant/signifié), *unité*, *signe*, *substitution* (de sens à signifié), *signifiant*, *Saussure*, etc. Outre cette isotopie, *sens* est significativement corrélé à *homme* (6.52, rang 2). *Sens* semble ainsi permettre d'articuler l'homme au signe, et par extension, à la langue et au langage.
- **Combettes** : *sens* est peu caractérisé, et d'ailleurs peu employé par Combettes. Les co-occurents du lexèmes sont globalement irréguliers ; au mieux relève-t-on un emploi récurrent de la tournure *aller dans le sens de...*
- **François** : *sens* est chez François quasi-exclusivement employé dans son texte *Sémantique* :



Graphique : Représentation de SENS chez François

- **Kerbrat** : on ne peut pas parler non plus d'isotopie ; *sens* ne semble pas débattu, et est corrélé aux unités observées par l'auteur (catégories : *mot(s)*, *unités* et exemples : *femelle*, *Sganarelle*, etc.).
- **Kleiber** : si *sens* était associé à un sème [+ instable] chez Authier, il est particulièrement stabilisé chez Kleiber, et quasi-exclusivement associé à des éléments partageant le sème [+stable] ou [+déterminé] :

(*sens*) *descriptif* (8.50) – qui est d'ailleurs quasi figé chez l'auteur, *détermine* (7.19), *conventionnel* (6.57), *dénominateur* (6.49), *stable* (6.16), *représentationnel* (6.02), *déterminé* (6.02), *conditions* (5.96), *psychologique* (5.96), *codé* (5.62), *stables* (5.36), *instructionnel* (5.36).

A noter, la présence de *sémantique* (8.41) comme cadre, spécifique à l'environnement paragraphe du lexème, et que l'on ne retrouve pas si l'on restreint le contexte d'apparition de la forme à 50 caractères.

- **Moirand** : le *sens* chez Moirand articule deux isotopies sociale et linguistique. On relève en effet pour la première les corrélats : *social(e)*, *histoire*, *mémoire*, *années*, de même que des items associés aux types de corpus socialement inscrits que décrit l'auteur : *manipulé*, *chargé*, (principe de) *précaution*, *historien*, *choses*, *devoir* (de mémoire), etc. ; et pour la seconde sèmes, *énoncés*, *sémantique*, *langagière*, *mots*.
- **Neveu** : outre le figement *sens lexical*, *sens* est globalement peu stabilisé dans les textes de Neveu dont nous disposons, comme le montre le nombre relativement restreint de co-occurent de la forme.
- **Rabatel** : peu de régularités significatives.
- **Rastier** : si *sens* et *signifié* s'interdéfinissaient chez Bergounioux, on le trouve aux côtés de *signification* chez Rastier (premier co-occurent relevé, écart de 15.59) – *sens* et

signification partagent d'ailleurs les mêmes corrélats, ce qui indique bien un emploi concomitant des deux notions. *Sens* s'inscrit ainsi dans un environnement sémiotique (*contenu, carré (sémiotique), signifié, signe*) et sémantique – on relève ainsi des éléments qui renvoient à différents modèles du sens (*contenu, mimesis, immanent, texte...*); bien qu'originellement distincts et susceptibles de polémiques, ces éléments sont conciliés dans le cadre sémantique rastiérien. Notons la présence significative de *texte*, palier empirique le plus élevé (3^e co-occurent de *sens*) et le fait que les deux éléments se sélectionnent mutuellement (resp. 2^e et 3^e premiers co-occurents, 5.83 et 7.74).

- **Siblot** : comme chez Barbéris, on observe une conception praxématique du sens (*production, praxème, récepteur, procédures, praxis, sociales, programmes, produire*) appliquée à l'objet *texte* (*texte(s), clôture*).

6.6.2. LANGUE

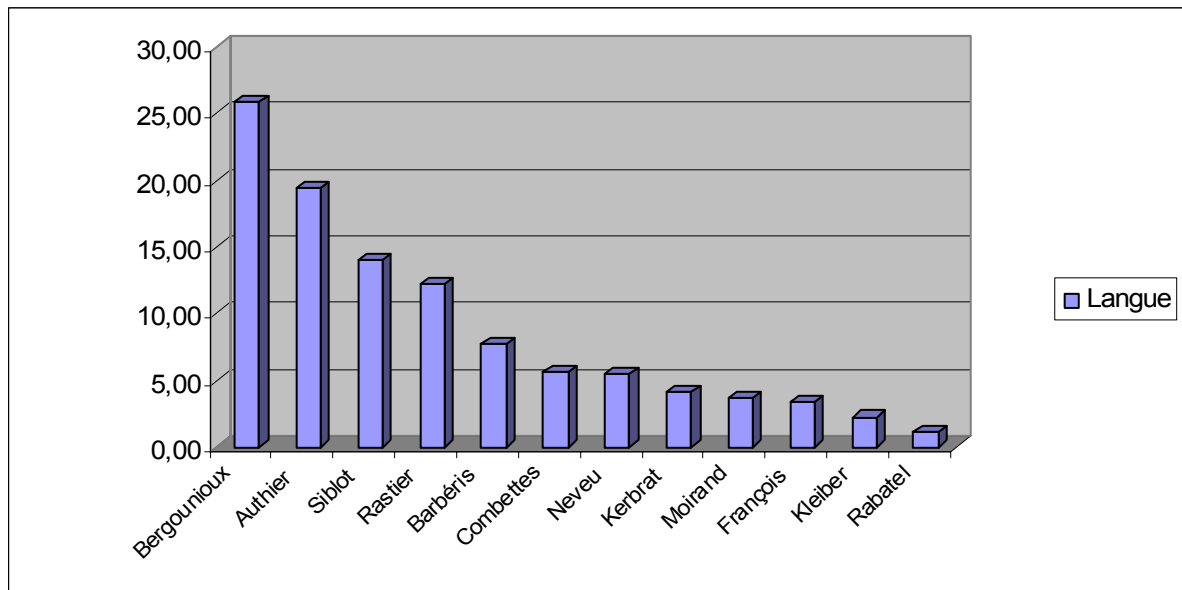
6.6.2.1. Différentiel ASL/Auteurs

Ecart	Corpus	Extrait	Mot
119.78	860	870	LANGUE
25.08	191	92	FRANÇAISE
15.18	327	83	SYSTÈME
14.47	25746	2032	LA
13.52	17	14	MATERNELLE
13.37	976	154	LINGUISTIQUE
12.45	39670	2875	DE
11.76	772	121	LANGAGE
11.42	1864	223	DISCOURS
10.80	11	9	PARLÉE
9.80	10658	853	L'
9.72	16	10	SGANARELLE
9.39	35	15	MORPHÈMES
9.10	6789	568	D'
8.82	74	22	SAUSSURE
8.67	35	14	ENDOPHASIE
8.56	89	24	NORMES
8.38	543	77	FRANÇAIS
8.35	37	14	USAGES
8.12	527	74	PAROLE
8.10	317	52	FAITS
7.97	22	10	POTENTIEL

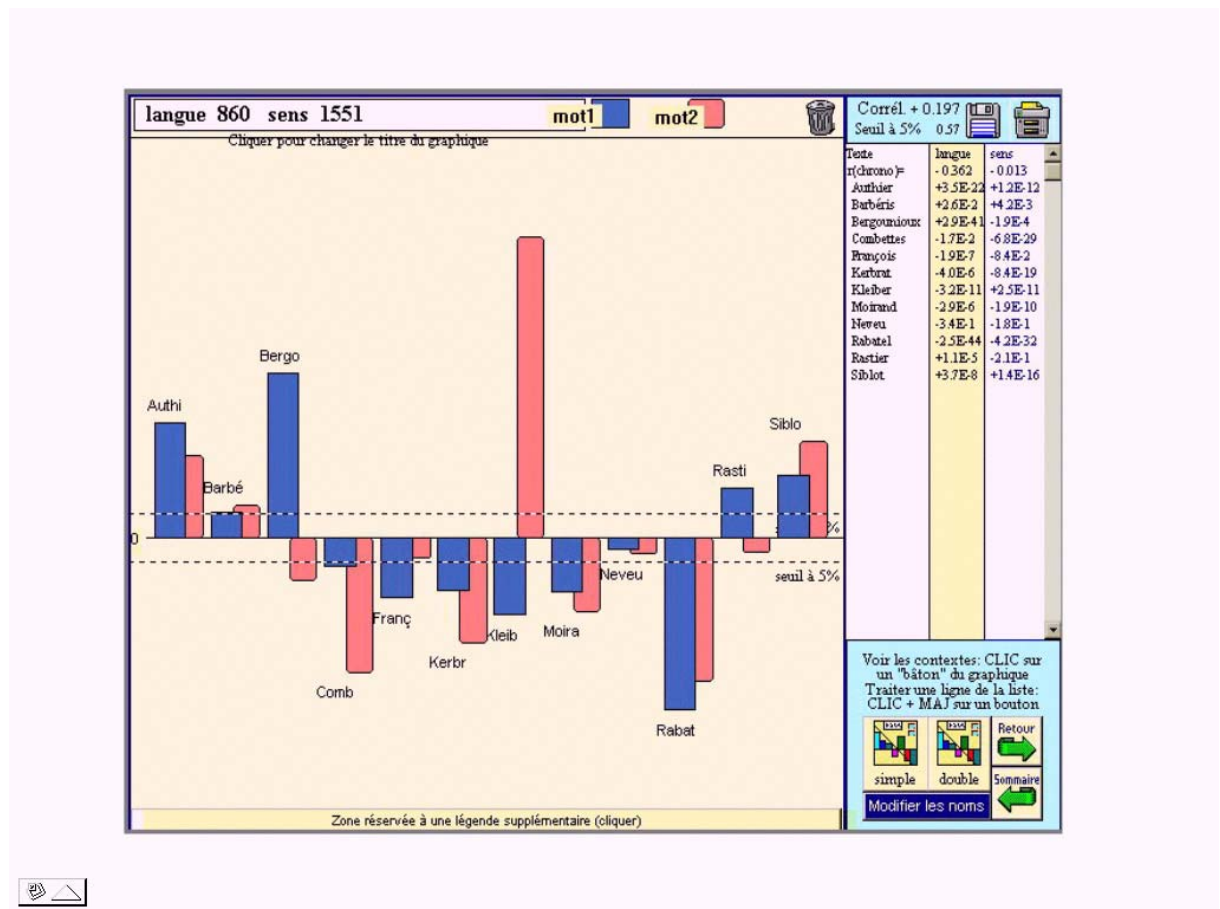
Tableau : Premiers co-occurents de LANGUE – Auteurs

Au regard des co-occurents obtenus sur le corpus *Auteurs* – et sur le seul numéro thématique *Contexte(s)-*, il semble que la *langue* soit invariablement associée à *linguistique*, *Saussure* et *système* et aux figements *langue française*, *langue maternelle* et *langue parlée*.

6.6.2.2. Quantitatif



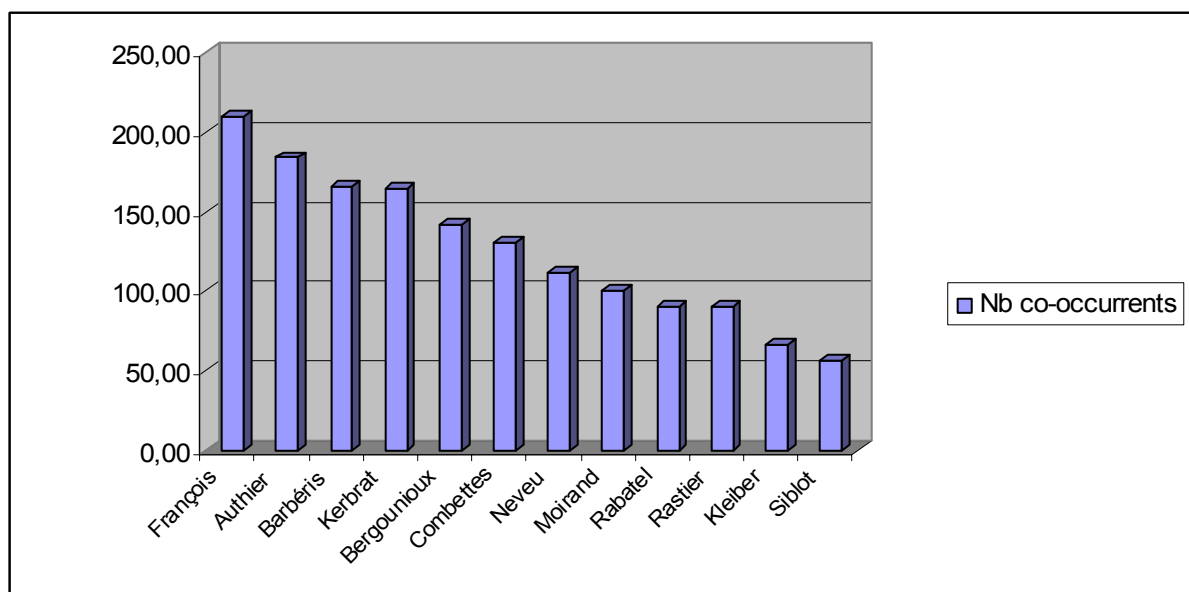
Graphique : Répartition de LANGUE (moyenne absolue par texte)



Graphique : Comparatif SENS/LANGUE – corpus Auteurs

On observe également de grandes disparités d'utilisation de *langue* d'un auteur à l'autre : Bergounioux y recourt ainsi 25 fois plus que Rabatel. Chez 9 auteurs sur 12, on note un emploi corrélé des deux notions de *langue* et de *sens* : Kleiber, que nous avons déjà observé comme recourant massivement à *sens*, est le seul auteur observé qui emploie *sens* au détriment de *langue*, tandis qu'on note l'inverse chez Bergounioux et Rastier.

♦ L'examen des co-occurents (graphique qui suit) montre que *langue* et *sens* sont peu, voire non stabilisés chez François (plus grand nombre de co-occurents observés malgré un recours faible aux deux objets), ce qui semble indiquer que ce sont des objets périphériques, voire peu discutés : *langue* et *sens* apparaissent d'ailleurs quasi-exclusivement dans le texte *Sémantique* de l'auteur. En revanche, les deux concepts sont fortement stabilisés chez Siblot et ne s'inscrivent que dans un fond sémantique ; il conviendrait de mettre au jour s'il s'agit d'une caractéristique stylistique ou domaniale (et spécifique au cadre praxématique) :



Graphique : Nombre de co-occurents de LANGUE

6.6.2.3. Analyse des co-occurents de langue

Langue drainant de nombreux lieutenants / figements (plus nombreux que pour *sens*), il n'est pas surprenant que les auteurs observés y recourent de manière caractéristique : *langue française* chez Bergounioux et Neveu, *langue maternelle* chez Authier, *langue parlée* chez Barbéris.

- **Authier** : on observe les figements *langue naturelle* et *langue maternelle*, de même que l'emploi de *système*.
- **Barbéris** : si Barbéris s'intéresse à la *langue parlée* (2^e co-occurent observé), cette dernière s'inscrit tout comme *sens* dans le cadre praxématique de l'auteur, d'où les co-occurents linguistiques et marxistes *classes*, *discours*, *légitime*, *dominante*, *souterraine*, *registres*, *populaire*, etc. On observe également un sème [+divers] dans les co-occurents *variété*, *spontanée*, *innombrables*, *variétés*, etc. (?)
- **Bergounioux** : perspective historique – outre la présence récurrente du figement *langue française*, liée à l'objet de l'auteur (histoire de la langue française au XIX^e), les co-occurents

observés ont des écarts faibles (moins de 5), et se rattachent à la perspective historique (*histoire, était, diachronique, etc.*)

- **Combettes** : la *langue* est nettement abordée d'un point de vue diachronique (*état(s), ancienne, anciens, catégories, textes, système, trace, surviennent, permettent, moderne, existence, faits, phénomènes, corpus, vestiges, etc.*).
- **François** : notion trop marginale pour être caractérisée.
- **Kerbrat** : on relève peu de régularités d'emploi et il est encore une fois délicat de parler d'isotopies ; les corrélats observés semblent plus ponctuels, et différent d'un texte, et d'un objet, à l'autre : opposition *masculin/féminin* dans le texte *Converser au féminin*, dialogue entre *Sganarelle* et *Panrace* lié à un article sur le malentendu. Si Kerbrat s'inscrit dans une perspective pragmatique, la diversité des objets abordés complexifie l'observation des concepts manipulés et la mise au jour de régularités significatives.
- **Kleiber** : il est difficile de dégager de réelles stabilisations de *langue*, le concept étant nettement moins central que *sens* chez l'auteur (18 vs. 404 occ.).
- **Moirand** : à l'instar de *sens*, la *langue* est socialement inscrite chez Moirand, et fortement corrélée à la notion de *discours*.
- **Neveu** : outre un emploi co-occurent de *langue* et de *linguistique*, et un figement *langue française* – d'ailleurs largement lié à plusieurs citations d'un numéro de la revue *Langue française* et de l'ouvrage de Rémy de Gourmont, *Esthétique de la langue française*, on relève peu de régularités significatives.
- **Rabatel** : *langue* apparaît chez Rabatel dans un fond essentiellement didactique (*élèves, maîtres, initiation, grammaire, appris, école, aider, etc.*).
- **Rastier** : si *sens* s'interdéfinissait avec *signification*, la *langue* est d'abord discutée aux côtés de la *parole* chez Rastier – Saussure (*saussure, saussurienne*) est d'ailleurs représenté ; on relève de surcroît les co-occurents *morphèmes* (le lexique des morphèmes comme inclus dans le système de la langue) et *normes* (la langue comme espace de normes). Niveau de normativité et de systémativité (*système, norme*), la langue s'oppose aux discours (*discursives, discours*), aux champs génériques (*champs, génériques*) et aux genres (*genre(s)*). On observe ici encore la stabilité du système conceptuel rastiérien,
- **Siblot** : si le *sens* s'inscrivait très nettement dans une conception praxématique, il en va quelque peu différemment de la *langue*, qui est discutée en termes sémiotiques saussuriens (*signes, Saussure, nomenclature, système, etc.*), et aux côtés du *langage* (*langage, langagières*). On soulignera que les co-occurents observés manifestent des écarts globalement faibles (5.4 max.).

6.6.3. Conclusions

La variable *Auteurs* a une incidence nette sur la thématization des concepts, ce qui n'est guère surprenant, les notions même les plus fréquentes étant constamment redéfinies et réappropriées.

Si la *langue*, qui s'avère être tactiquement le concept le plus problématisé, semble plus stabilisée que le *sens*, concept le plus employé, cette dernière notion semble constituer un meilleur point d'entrée dans les systèmes conceptuels développés par les auteurs observés : cette distinction pourrait éventuellement participer à celle de concepts de fond / concepts de forme proposée par (Rastier, 2003). *Langue* serait ainsi un concept de fond disciplinaire peu débattu tandis que *sens*, qui paraît moins stabilisé, serait une forme plus discutée.

De manière générale, on peut opposer les concepts dont les co-occurents sont d'autres concepts des entrées dont les co-occurents sont des exemples ou des objets de description linguistique : dans le premier cas, le concept semble (inter)défini, voire peut-être débattu, tandis que dans le second, il semble essentiellement instrumental ou méthodologique.

Si on ne considère que les concepts interdéfinis, on peut *a fortiori* opposer les concepts inscrits dans des cadres théoriques et méthodologiques (e.g. Praxématique, Sémantique Interprétative, Sémantique de la référence klébérienne, et cadre énonciatif d'Authier) des concepts ponctuellement débattus où le concept reste de faible fréquence, ou n'est pas pivot dans un système conceptuel – e.g. cadre historique de Bergounioux (sens/signe). Dans le premier cas, on a pu distinguer pour *sens* le cadre fondationnel philosophique de la praxématique (vocabulaire marxiste, gestaltiste, etc.) et le cadre épistémologique de Rastier, où il est le lieu de mises en relation de différentes traditions culturelles (sémiotique, héritage philosophique, herméneutique, structuralisme, etc.).

6.7. Conclusion

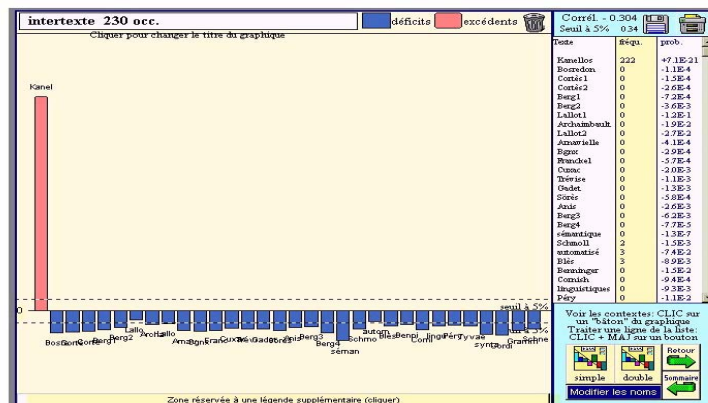
Comme nous l'avions annoncé au départ, le présent chapitre est finalement plus exploratoire que descriptif : nous avons tenté d'évaluer l'intérêt et la pertinence descriptives de six critères typologiques. L'entreprise mériterait naturellement d'être approfondie et étendue à l'ensemble des concepts de notre corpus, dans la mesure où nous sommes particulièrement intéressés aux concepts de haute fréquence *sens* et *langue*.

Si les substantifs de haute fréquence ébauchent le fond disciplinaire de la linguistique, leur seule prise en compte est insuffisante, dans la mesure où leur statut thématique demeure peu défini : s'agit-il de concepts discutés, instrumentaux ou encore des lexicalisations lieutenantes d'une forme ?

L'examen des corrélats morphosyntaxiques et lexicaux des formes permet en partie de répondre à cette interrogation en discriminant différentes acceptions du concept. Si les co-occurents lexicaux permettent de distinguer, et d'isoler les formants, les corrélats morphosyntaxiques semblent particulièrement discriminants – à condition d'être comme nous au fait de la morphosyntaxe du corpus : ainsi, *sens* au singulier renverrait bien à une lexicalisation privilégiée de concept, tandis que *les sens* seraient des formes peu discutées, voire instrumentales car corrélées aux marqueurs de l'exemple.

La disposition tactique du candidat dans l'article permet de mettre au jour de manière très claire les formes discutées des formes non débattues, ce qui est particulièrement intéressant, dans la mesure où le critère est généralement peu pris en compte : le logiciel CR de S. Loiseau qui permet de telles manipulations de corpus nous semble ainsi promis à un bel avenir.

Si les critères pris en compte se sont avérés présenter un intérêt descriptif et discriminant, la liste est bien entendu ouverte : nous avons par exemple écarté le critère syntaxique. De surcroît, si l'on admet que les concepts les plus fréquents et les plus stabilisés font partie du fond disciplinaire (Rastier, 2005), les concepts émergents devraient logiquement être plus discutés, donc moins stabilisés ; il en va ainsi par exemple du concept d'*intertexte*, inconnu du glossaire bibliographique Gobert, mais de fréquence suffisante pour se prêter à l'analyse statistique – le numéro thématique 1 du corpus lui est en effet dédié :



Graphique : Représentation du concept d'intertexte dans les 32 numéros de revue

Enfin, nous avons pu observer une incidence très forte du numéro thématique de revue, et de l'auteur, dont les poids respectifs doivent donc être précisément évalués.

Le chapitre suivant est ainsi dédié au style d'auteur ; nous entrons donc dans la seconde phase de notre entreprise descriptive, en confrontant le corpus à d'autres collections, et à d'autres typologies.